Article

A Similarity-Based Hierarchical Clustering Method for Manufacturing Process Models

Hyun Ahn [1] and **Tai-Woo Chang** [2,*]

1   Division of Computer Science and Engineering, Kyonggi University, Suwon, Gyeonggi 16227, Korea; hahn@kgu.ac.kr
2   Department of Industrial and Management Engineering/Intelligence and Manufacturing Research Center, Kyonggi University, Suwon, Gyeonggi 16227, Korea
*   Correspondence: keenbee@kgu.ac.kr; Tel.: +82-31-249-9754

check for updates

Abstract: As the adoption of information technologies increases in the manufacturing industry, manufacturing companies should efficiently manage their data and manufacturing processes in order to enhance their manufacturing competency. Because smart factories acquire processing data from connected machines, the business process management (BPM) approach can enrich the capability of manufacturing operations management. Manufacturing companies could benefit from the well-defined methodologies and process-centric engineering practices of this BPM approach for optimizing their manufacturing processes. Based on the approach, this paper proposes a similarity-based hierarchical clustering method for manufacturing processes. To this end, first we describe process modeling based on the BPM-compliant standard so that the manufacturing processes can be controlled by BPM systems. Second, we present similarity measures for manufacturing process models that serve as a criterion for the hierarchical clustering. Then, we formulate the hierarchical clustering problem and describe an agglomerative clustering algorithm using the measured similarities. Our contribution is considered on the assumption that a manufacturing company adopts the BPM approach and it operates various manufacturing processes. We expect that our method enables manufacturing companies to design and manage a vast amount of manufacturing processes at a coarser level, and it also can be applied to various process (re)engineering problems.

Keywords: manufacturing process model; business process management; hierarchical clustering; similarity; BPMN

1. Introduction

Thanks to the evolution of manufacturing systems, manufacturing companies can efficiently plan, design, and produce their products. However, as the adoption of the internet of things (IoT) and manufacturing information systems increases, the complexity and interactivity of communications within information systems is expected to expand. Furthermore, there may be large volumes of data stored in a database, because IoT devices constantly transmit data of identified objects [1]. Because of these circumstances, it is still challenging for manufacturing companies to manage a vast array of their manufacturing data and processes. Although information and communication technologies have been introduced to promote technical support for manufacturing operations management (MOM), a more comprehensive methodology should be adopted to fully support manufacturing-process-centric management activities.

The business process management (BPM) approach can be a solution to tackle this hurdle by continually improving processes through automation and optimization. Manufacturing companies could also benefit from the well-defined methodologies and process-centric engineering practices

of the BPM approach for optimizing their manufacturing processes. In particular, various analysis techniques which come from the BPM research field may be fruitful for efficiently managing a large set of manufacturing processes [2].

Based on the BPM approach, we propose a similarity-based hierarchical clustering method for manufacturing process models to facilitate managerial activities of manufacturing processes at a group level (e.g., the design and engineering tasks). To this end, modeling of manufacturing processes and enabling the measuring similarities must be performed beforehand. Therefore, this paper exploits the business process model and notation (BPMN) standard and its extension for modeling of manufacturing processes. The novel manufacturing process of modeling with the adaptation of the BPMN helps human understanding by a visual diagram and machine understanding is helped by a converted textual form that is executable in the BPM system. Measuring similarities, in this study, is performed by calculating the operation similarity and structural similarity between the models of manufacturing processes, respectively. The hierarchical clustering method this paper proposes operates based on the measured similarities between the models. We note that the contributions of this paper assume that a manufacturing company adopts the BPM approach and it operates a variety of manufacturing process models on a scale that is difficult to manage manually.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce previous works related to this work. Next, in Section 3, we describe the adaptation of the BPM approach to the MOM area and its primary activities derived from this combined environment. Section 4 includes details of BPMN-based manufacturing process modeling. As the main work of this paper, Section 5 presents a similarity method for BPMN-based manufacturing process models, and Section 6 describes a hierarchical clustering phase of manufacturing process models based on the similarities we propose. Finally, this work will be concluded with future research in Section 7.

2. Related Works

The main elements of the method we propose in this paper are the similarity measures and clustering based on such similarity. We, therefore, organize the content of this section into two parts: process model similarity and clustering. Due to the lack of studies addressing similarity for manufacturing processes, a series of studies that propose similarity measures for business process models are mainly introduced and summarized in the similarity part. On the other hand, the clustering part includes related works applying clustering methods to the field of production research, and a comparison between them and this study.

Many similarity measures have been proposed in the field of BPM to handle a large collection of process models. Jung et al. [3] presented a similarity measure which is based on the execution probability and structural features of business process models. Dijkman et al. [4] investigated the similarity problem of process models centered on the structural aspect and proposed graph-matching algorithms for measuring similarities. They considered business process models as graphs and defined the structural feature between graphs, namely the graph-edit distance, as the criterion of similarity. Many different similarity concepts have been proposed from other studies, and a few comprehensive studies suggested that similarity measures between business process models can support the search and reuse of similar process models [5–7]. Despite a rich set of previous studies contributing to the process similarity research, there is still a lack of proper methods for manufacturing processes since the BPM approach has not been actively discussed in manufacturing industry. As slightly different applications, in [8], the similarity concept was applied to the problems of machine groupings for the design of manufacturing systems. Compared to our similarity, this study focused on relations between machines and components and therefore such measures are not process-centric.

On the other hand, clustering methods have been applied in manufacturing to address various manufacturing engineering problems including the machine grouping [9,10], CAD model grouping [11], product variants management [12], and product release planning [13]. The machine grouping problem has traditionally been the subject of manufacturing systems, especially cellular manufacturing.

Dimopoulos and Mort [9] presented the machine-grouping technique based on the combination of genetic programming and hierarchical clustering. Likewise, Park and Suresh [10] addressed the part-machine grouping to identify families of parts with similar routing sequences. The hierarchical clustering method, considered as a solution of the grouping problem, is supported by using the Fuzzy ART neural network. Their experiments showed the compelling performance of hierarchical clustering for large data sets. These two studies provided prerequisites for the implementation of cellular manufacturing. As a different kind of grouping problem, Li and Xie [11] proposed a module partition approach with the hierarchical clustering of components to group CAD models into modules based on component dependencies. In addition, hierarchical clustering methods have also been applied to engineering applications, such as product variety management [12] and the optimization problem of deciding the contents of product releases [13].

As described above, hierarchical clustering methods have been applied for grouping elements that are part of a manufacturing process—such as machines, components, and CAD models—or to engineering applications that are irrelevant to the process grouping. This paper tries to combine the BPM approach to the field of MOM and proposes a hierarchical clustering method based on the similarity for grouping manufacturing process models, eventually enabling process group-level operations and engineering applications. Therefore, we believe that our work is distinguished from the previous works mentioned above.

3. Applying the BPM Approach to Manufacturing Operations Management

MOM is a methodology with the aim of optimizing the manufacturing operations in the stages including the creation, planning, production, and distribution of products. Because of the significance of each operation stage, there are well-established MOM systems (e.g., manufacturing execution systems) that have been developed to provide systematic support to the stage-level operations. In general, MOM systems are broadly classified into the following categories: production management system, performance analysis system, quality and compliance system, and human–machine interface system.

Despite the merits from such MOM methodology and systems, we believe that there is still a necessity to apply a process-centered approach to the manufacturing processes to concentrate the optimization efforts. Manufacturing companies have their own manufacturing processes as their core assets. Therefore, the management capability of such processes has a significant impact on the performance of manufacturing to a greater or lesser extent. It emphasizes that the collaboration capability supporting people-to-people, people-to-systems, and systems-to-systems interactions in performing manufacturing operations should be incorporated in order to meet technological requirements that emerge from trends in MOM systems.

The BPM approach [14] covers management methodologies and systems for business processes. However, it can also be an effective way to integrate and align the capabilities of manufacturing operations with the manufacturing processes. While existing MOM systems are intended to increase the efficiency of operations associated with the product lifecycle, BPM systems can be a valuable addition to MOM since they offer comprehensive capabilities for manufacturing process management. Therefore, manufacturing companies can achieve advantages of the BPM approach through the modeling and efficient implementation of their manufacturing processes.

The process model is a central concept of BPM, formed from the contents created during the process design, and can be executed automatically by the process engine. Relating to the manufacturing processes, Figure 1 represents details of the activities and expected outcomes based on manufacturing process models that correspond to each stage consisting in the BPM lifecycle.

- Design Stage—This stage includes a series of steps from the establishment of initial development plan for the product to the redesign of existing manufacturing processes. From the view of process, this stage specifies which manufacturing operations should be taken to produce the desired product, and which components are required for each operation. Basic production-related

data—such as bill-of-materials (BOM), operation process charts, and other specifications—are the expected outcomes of this stage and will serve the modeling stage of manufacturing processes.

- Modeling Stage—Through the modeling of manufacturing processes, we can shift the manufacturing processes to the management area of BPM and ensure their automated execution. The outstanding standards for modeling processes, such as the BPMN and BPEL (Business Process Execution Language), can be applied in order to make manufacturing processes into the form of standard-compliant and executable models.

- Execution Stage—Modeled manufacturing processes can be automatically executed by BPM systems under the assumption that manufacturing execution and control capabilities (e.g., manufacturing execution systems) are integrated or communicated with the BPM system. Each task of a manufacturing process model implements an individual manufacturing operation on the shop floor. During executions of manufacturing processes, event log data containing detailed records of manufacturing operations are generated and stored. Collected event log data are used in various analytics stages (e.g., process monitoring and optimization).

- Monitoring Stage—As event logs are generated while manufacturing processes are being executed, it is possible to monitor the status of running manufacturing processes or process instances in real time [15]. The well-established data formats for event logging can support the monitoring stage. For example, the XES (eXtensible Event Stream) standard, which has been widely adopted in the BPM field, has extensible data schema to incorporate specialized attributes of processes from various industry domains. As the result of this stage, manufacturing process reports help process administrators to take appropriate actions.

- Optimization Stage—Based on event logs, process mining [16,17] and post-analysis techniques provide us with opportunities to investigate the quality and performance of completed manufacturing process executions and related operations and improve those inefficient parts. The derived requirements for manufacturing process improvement will be referenced in the process design or redesign stage of the next cycle.
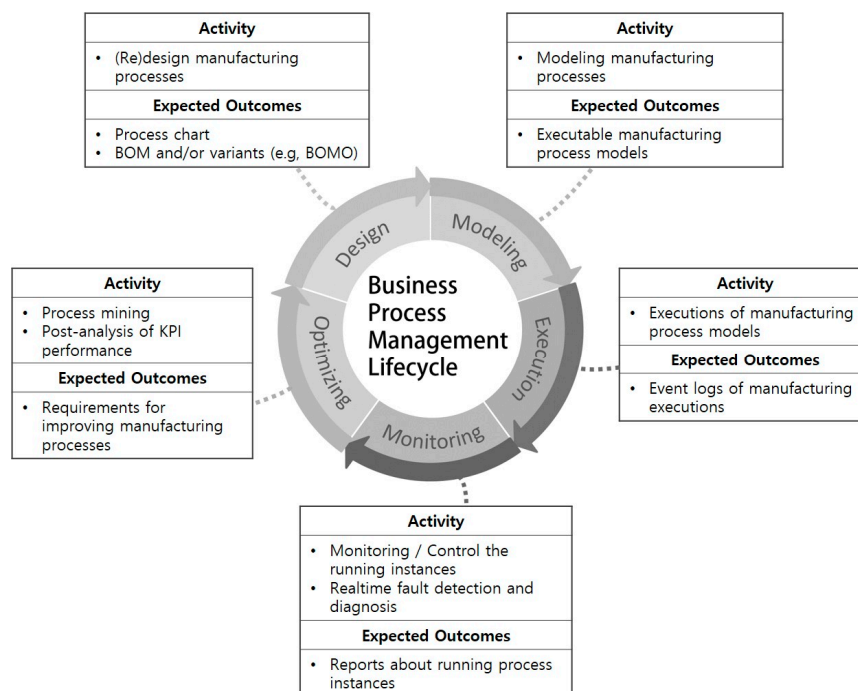
Figure 1. BPM lifecycle including the core stages in managing processes and details of activities and outcomes expected from the adoption of BPM into the MOM area.

Through continuous repetition of the BPM lifecycle applied into the MOM, manufacturing processes aligned with this management cycle can be expected to be constantly optimized to reflect the requirements that are derived by internal and external management factors. To take advantage of the convergence of these two management methodologies, modeling of the manufacturing process is a mandatory task. Therefore, the next section describes how to model the manufacturing process so that it complies with BPM-related standards from document-based manufacturing process data.

4. BPMN-Based Manufacturing Process Modeling

BPMN is one of the most widely used standards for modeling business processes. It provides a rich set of element types that can fully represent the context of a business process. Moreover, this standard can be easily extended, and it has been applied to modeling problems in various domains, such as wireless sensor network [18], healthcare processes [19], and manufacturing processes [20,21]. This section describes a construction procedure for manufacturing process models from the fundamental data of manufacturing processes.

Manufacturing companies conventionally possess operation process charts and BOM specifications to define and manage their manufacturing processes. In the Republic of Korea, the KS-A-3002 standard [22], which is for manufacturing process charts, has been used in manufacturing industries. However, it provides only a set of graphical notations, and there is no technical support for the modeling and automatic executions for manufacturing processes. Figure 2 shows an example of a manufacturing process chart for a thermocouple probe product.

A BOM specification is an essence of defining a manufacturing process that contains detailed data about the components (e.g., raw materials, parts, subassemblies, and end products) of each needed to manufacture a certain product. However, it is not sufficient to express the production flow of the manufacturing process. A manufacturing process consists of a set of manufacturing operations, which require a collection of components to manufacture and have prior relationships with other preceding and/or successive operations. In this regard, we additionally exploit the BOMO (Bill of Material Operations) concept, which is presented in [23] to define production-flow-oriented information of manufacturing process examples.

As shown in Table 1, each operation, such as wire welding, consumes a set of components and produces an intermediate component or end product. The preceding operation information provides an execution ordering of the operations within a manufacturing process. For example, the wire welding (OP1) and quartz tube winding (OP2) operations can be performed in parallel, but they must precede the wire injection operation (OP3) according to their ordering relationships. Through these basic ingredients of manufacturing process data (manufacturing process chart, BOM, and BOMO), we can organize structures for manufacturing processes.

Table 1. BOMO data of the product examples [2].

End Product	Operation		Component	Intermediate Component	Preceding Operation
	ID	**name**			
PROBE-01	OP6	Packaging	SUB-05, PACK-01, PACK-04, PACK-06, PACK-13, PACK-17	PROBE-01	OP5
	OP5	Insulation (cement)	SUB-04, MATL-11	SUB-05	OP4
	OP4	Housing	SUB-03, PART-08, PART-18, PART-19	SUB-04	OP3
	OP3	Wire injection	SUB-01, SUB-02	SUB-03	OP1, OP2
	OP2	Quartz tube winding	PART-15, MATL-02, MATL-04	SUB-02	-
	OP1	Wire welding	MATL-03, MATL-05, MATL-06	SUB-01	-

To create manufacturing process models, we apply the BPMN standard, with an extension of notations. The BPMN standard has a variety of its extensions, but it lacks modeling notations for the manufacturing domain. Although a few studies presented extensions for manufacturing processes [20,21], their extensions do not cover the full context of the manufacturing domain due to

the absence of uniformity. Accordingly, we define a BPMN extension that has a minimal set of element types but suffices to model the examples we present in this paper.
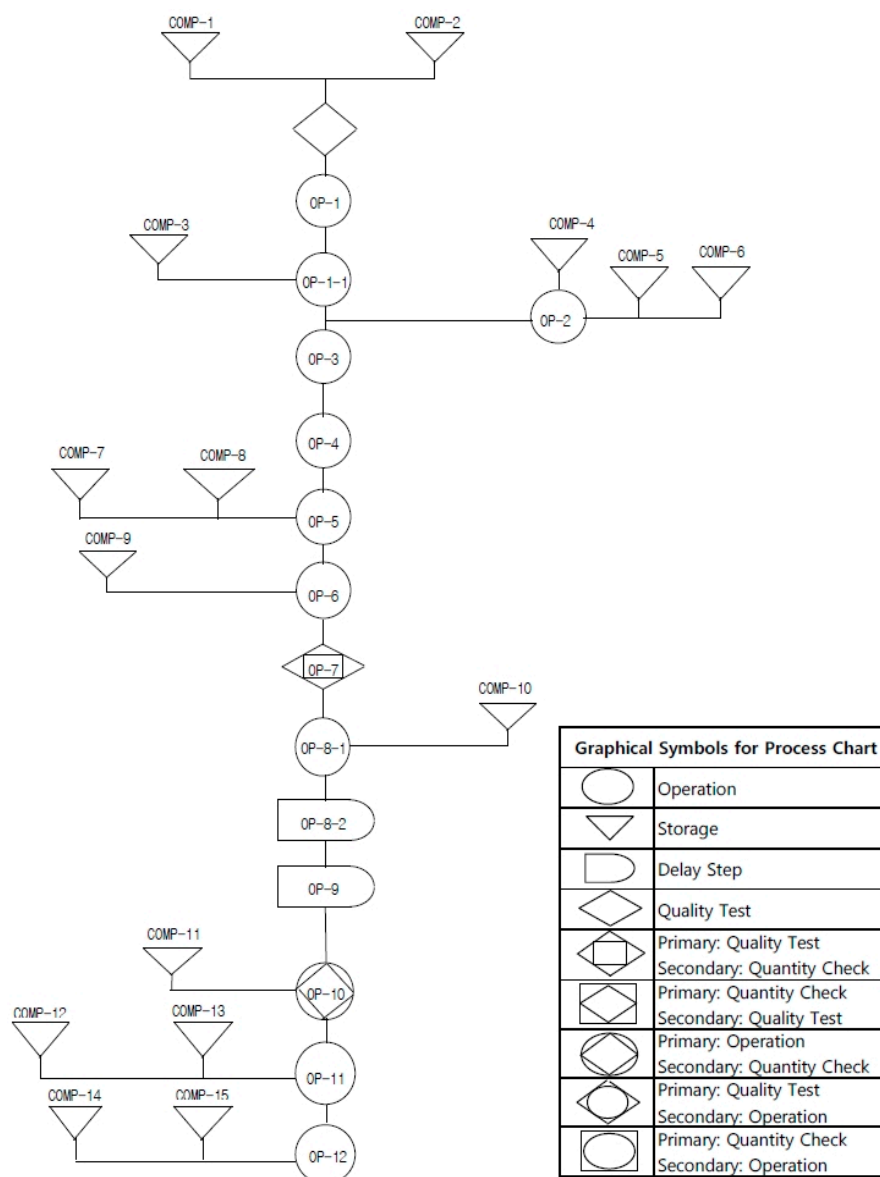

Figure 2. Operation process chart for a thermocouple probe product.

Every manufacturing process model essentially contains a start event and end event indicating their starting/terminating points. An operation refers to a primitive work within a manufacturing process. This type has precedence and/or successive relationships with other operations through connecting objects of sequence flows. Also, it is related to particular components through connecting objects of component associations, and production-related information such as operating cost. A component refers to an input of a particular operation and it can be a raw material, part, and subassembly.

Regarding the control-flow aspect, the notations of Table 2 are limited to focus on the examples including only sequential and parallel control-flow patterns. However, it is necessary to add extra notations for other patterns, such as selective, repetitive, and other complex patterns to facilitate modeling of advanced types of manufacturing processes and systems (e.g., flexible manufacturing systems and reconfigurable manufacturing systems). Figure 3 shows a result of the BPMN manufacturing process modeling for the thermocouple product specified as Table 1.

Table 2. BPMN notations for manufacturing process models [2]

Notation	Element type	Description
◯	Start event	A 'start event' indicates where a particular manufacturing process will start.
◉	End event	An 'end event' indicates where a manufacturing process will terminate.
▱	Operation	An 'operation' is a generic term for manufacturing tasks. Each 'operation' can be performed by machines and/or human workers.
▱	Component	A 'component' is a generic term for raw materials, assemblies, and parts needed to manufacture a product.
✛	Parallel gateway	A 'parallel gateway' is used to create and synchronize disjunctive flows which proceed in parallel fashion.
⟶	Sequence flow	A 'sequence flow' is used to show the order that operations will be performed in a manufacturing process.
┈┈▶	Component association	A 'component association' is used to link components (e.g., material, part) and operations.
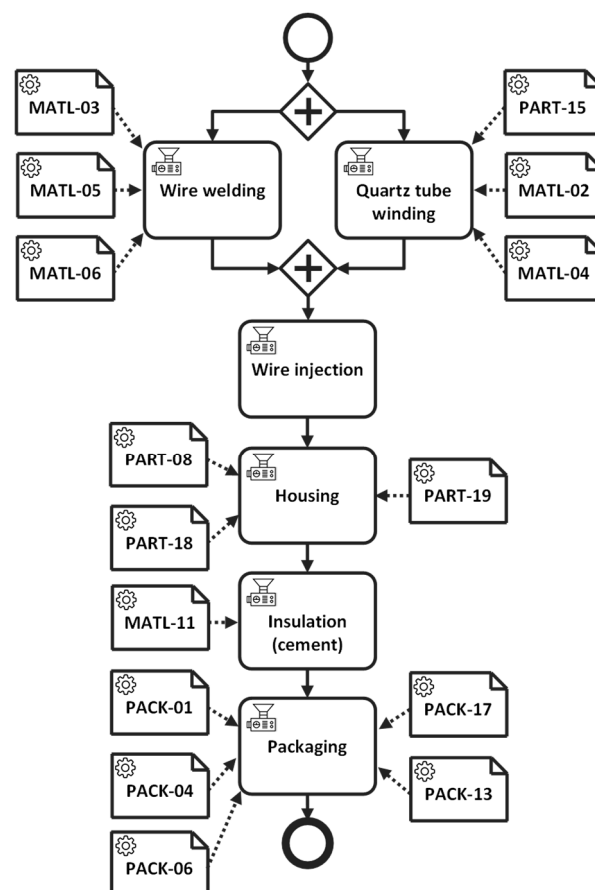
Figure 3. An exemplary model of a transformed BPMN manufacturing process.

The BPMN model is represented by a visual diagram (see Figure 3), while at the same time it is converted to a textual form (see Figure 4) that is executable by BPM systems. Figure 4 represents fragments of the textual content of the manufacturing process model example. This content represents detailed information, including definitions of operations, components, and connecting

objects (e.g., gateway, sequence flow, and component association). When the manufacturing process model is instantiated, the execution engine of the BPM system interprets the information and creates a process instance to execute the manufacturing process with full control.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<bpmn:definitions xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:bpmn="http://www.omg.org/spec/BPMN/20100524/MODEL">
  <bpmn:process id="_q1GEAP5u" name="BT Probe">
    <bpmn:componentSet id="btprobe_componentSet">
      <bpmn:component id="_q1xZcP5u" name="MATL-03"/>
      <bpmn:component id="_q3EZ8P5u" name="MATL-05"/>
      <bpmn:component id="_q5g7YP5u" name="MATL-06"/>
                       ⋮
    </bpmn:componentSet>
    <bpmn:startEvent id="_q2AqAP5u" name="Start BT probe process"/>
    <bpmn:parallelGateway id="_q4J_EP5u" name="AND-Split Gateway-01"/>
    <bpmn:operation id="_q2VaIP5u" name="Wire welding">
      <bpmn:documentation>Cutting and welding a platinum wire.</bpmn:documentation>
    </bpmn:operation>
    <bpmn:operation id="_q4ogMP5u" name="Quartz tube winding">
      <bpmn:documentation>Bending a quartz tube using oxygen and LPG.</bpmn:documentation>
    </bpmn:operation>
    <bpmn:parallelGateway id="_q_9QMP5u" name="AND-Join Gateway-01"/>
    <bpmn:operation id="_q4eIIP5u" name="Wire injection">
      <bpmn:documentation>Injecting the welded wire into the quartz tube.</bpmn:documentation>
    </bpmn:operation>
                       ⋮
    <bpmn:endEvent id="_q4UXIP5u" name="Terminate BT probe process"/>
    <bpmn:sequenceFlow id="_rAI2YP5u" name="OP1-to-G01" sourceRef="_q2VaIP5u" targetRef="_q_9QMP5u"/>
    <bpmn:sequenceFlow id="_rAUckP5u" name="OP2-to-G01" sourceRef="_q4ogMP5u" targetRef="_q_9QMP5u"/>
    <bpmn:sequenceFlow id="_rAUckP5u" name="G01-to-OP3" sourceRef="_q_9QMP5u" targetRef="_q4eIIP5u"/>
                       ⋮
    <bpmn:componentAssociation id="_rBAZEP5u" name="MATL-03-to-OP1"
    sourceRef="_q1xZcP5u" targetRef="_q2VaIP5u"/>
    <bpmn:componentAssociation id="_rBN0cP5u" name="MATL-05-to-OP1"
    sourceRef="_q3EZ8P5u" targetRef="_q2VaIP5u"/>
    <bpmn:componentAssociation id="_rBEC2P5u" name="MATL-06-to-OP1"
    sourceRef="_q5g7YP5u" targetRef="_q2VaIP5u"/>
                       ⋮
  </bpmn:process>
</bpmn:definitions>
```

Figure 4. An example of textual representation of the BPMN manufacturing process model.

5. Similarity Measure

Modeled manufacturing processes can be efficiently executed, tracked, and analyzed by the abundant toolsets in the BPM field. In this section, we describe similarity measures for manufacturing process models that serve as a basis of hierarchical clustering. Specifically, in order to capture the characteristics of manufacturing process, we define and explain two sub-types of similarity: operation similarity and structural similarity.

5.1. Preliminaries

We denote a set of manufacturing process models as $\mathbf{M} = M_1, \ldots, M_n$ with n indicating the number of manufacturing process models, a set of operations as $\mathbf{OP} = OP_1, \ldots, OP_m$ with m indicating the number of operations, and a set of components as $\mathbf{C} = C_1, \ldots, C_l$ with l indicating the number of components. We also introduce a mapping function $\delta(\mathbf{OP}) \rightarrow \wp(\mathbf{C})$ that maps from an operation OP_k^i to input components of OP_k^i which are a subset of total components, where $OP_k^i \in \mathbf{OP}$ is the operation OP_k in the process model M_i.

5.2. Operation Similarity

Although many similarity measures have been proposed for business processes, the manufacturing process has distinguishing features from business processes. In particular, production-related

factors—which determine the characteristics of a manufacturing process—must be taken into account in measuring similarities.

The operation similarity we introduce in this paper is a similarity concept based on associations between operations and components that is one of these influential factors. Mostly, a manufacturing operation consumes a group of certain components to produce end products or intermediate components. Based on this feature, operation similarities between operations of the same type in different processes can be quantified. If two operations are the same type but have associations with different set of component types, these two operations are considered to have different characteristics. Mathematically, this similarity is based on the Jaccard coefficient, which is calculated by the division of the number of elements in the intersection set by the number of elements in the union set. Accordingly, the operation similarity between each operation OP_k in the process models M_i and M_j is calculated by Equation (1).

$$J\left(OP_k^i, OP_k^j\right) = \frac{\left|\delta\left(OP_k^i\right) \cap \delta\left(OP_k^j\right)\right|}{\left|\delta\left(OP_k^i\right) \cup \delta\left(OP_k^j\right)\right|} \tag{1}$$

For example, there are two 'Packaging' operations (OP_6) of the same type that is included in two different process models M_1 and M_2 as shown in Figure 5. The 'Packaging' operation in M_1 (Figure 5a) is associated with the set of input components $\delta\left(OP_6^1\right)$ = {PACK-01, PACK-04, PACK-06, PACK-13, PACK-17}, that is different from the set of input components of the 'Packaging' operation in M_2 (Figure 5b), $\delta\left(OP_6^2\right)$ = {PACK-01, PACK-03, PACK-10, PACK-14, PACK-17}. Accordingly, the operation similarity between these two operations is $J\left(OP_6^1, OP_6^2\right) = \frac{2}{8} = 0.25$.
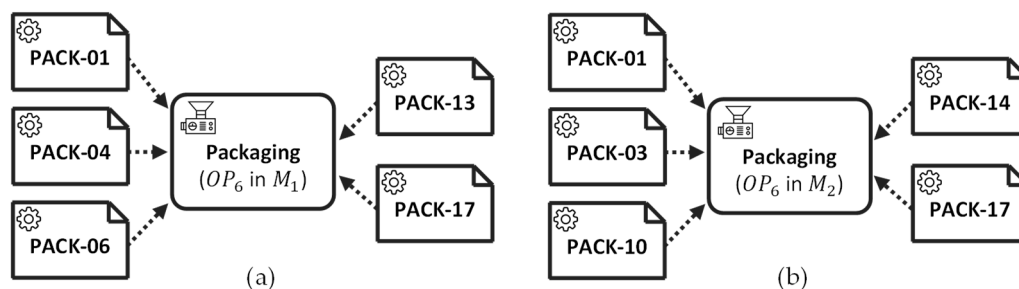

Figure 5. Operation–component associations.

We can measure operation similarities of the example models by calculating Jaccard coefficients based on the component association relationships of common operations (OP_1, OP_2, OP_3, OP_4, OP_5, and OP_6). In this regard, an operation similarity matrix including operation similarity measurements for all pairs of process models is defined as $X = (x_{rk}) \in \mathbb{R}^{\frac{n(n-1)}{2} \times m}$, where an element x_{rk} represents an operation similarity measure $J\left(OP_k^i, OP_k^j\right)$ between two operations of OP_k for the rth process model pair of the process models M_i and M_j. The index of the process model pair r is calculated by $r = \left(\sum_{a=1}^{i} n - 1\right) - (n - j)$, where $i < j$.

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{\frac{n(n-1)}{2}, 1} & \cdots & x_{\frac{n(n-1)}{2}, m} \end{pmatrix} \tag{2}$$

Based on the above, let M_1 and M_2 be two process models as specified in Table 3. The operation similarity matrix of the example equals to the row vector represented by X = [0.67, 1, 0, 0.40, 0, 0.25, ..., 0] since the example contains only two process models. Measured operation similarities affect the total similarity between two process models.

Table 3. Example of two process models including operation–component associations

Process Model	Operation	Components (without subassemblies)
Model 1 (M_1)	OP_1 (wire welding)	MATL-03, MATL-05, MATL-06
	OP_2 (quartz tube winding)	PART-15, MATL-02, MATL-04
	OP_3 (wire injection)	-
	OP_4 (housing)	PART-08, PART-18, PART-19
	OP_5 (insulation)	MATL-11
	OP_6 (packaging)	PACK-01, PACK-04, PACK-06, PACK-13, PACK-17
Model 2 (M_2)	OP_1 (wire welding)	MATL-05, MATL-06
	OP_2 (quartz tube winding)	PART-15, MATL-02, MATL-04
	OP_3 (wire injection)	-
	OP_4 (housing)	PART-02, PART-04, PART-18, PART-19
	OP_5 (insulation)	MATL-09, MATL-10, MATL-13
	OP_6 (packaging)	PACK-01, PACK-03, PACK-10, PACK-14, PACK-17
	OP_7 (fastening)	PART-10, MATL-01, MATL-07
	OP_8 (bar welding)	PART-09, PART-19
	OP_9 (sealing)	PACK-02, PACK-05

5.3. Structural Similarity

In this paper, we employ the similarity concepts of the activity vector and transition vector similarities, both are presented in [3]. We redefine these similarity concepts to fit manufacturing process models and call them structural similarity. As compared to the operation similarity based on the associations between operations and components, the structural similarity focuses on the existence of operations and the control dependency between operations. The structural similarity has two parts: operation vector similarity and transition vector similarity, and these concepts are slightly different from the similarity concepts for business process models in [3]. The total similarity is measured by putting these two similarities together.

5.3.1. Operation Vector Similarity

A manufacturing process comprises multiple operations that consume certain components and produce intermediate components or end products. Therefore, information indicating whether a specific operation is included in the process is the salient feature that characterizes manufacturing processes in terms of structural aspect. An operation vector v_i^O is an m-dimensional vector, where each element $v_{k,i}^O$ is a binary value (0 or 1) indicating whether the operation OP_k is included in the process model M_i.

$$v_i^O = \left[v_{1,i}^O, \ v_{2,i}^O, \ \ldots, \ v_{m,i}^O \right] \tag{3}$$

For example, Figure 6 represents two operation vectors v_1^O and v_2^O corresponding process models M_1 and M_2.
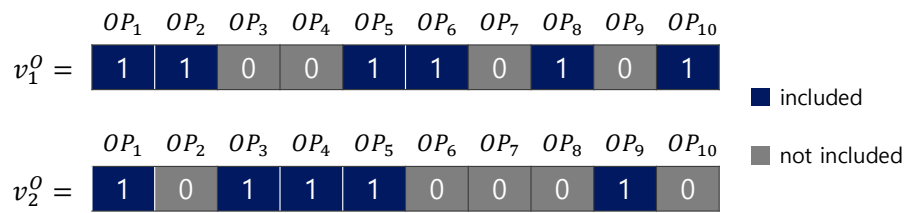
Figure 6. Representation of two operation vectors.

The operation vector similarity $sim_{ov}(M_i, M_j)$ is measured based on the Cosine coefficient and operation similarity. Given two operation vectors corresponding to two different manufacturing process models respectively (M_i and M_j), the following equation quantifies the similarity between these two operation vectors.

$$sim_{ov}(M_i, M_j) = \frac{\sum_{k=1}^{m} v_{k,i}^{O} v_{k,j}^{O} x_{rk}}{\sqrt{\sum_{k=1}^{m} \left(v_{k,i}^{O}\right)^2} \sqrt{\sum_{k=1}^{m} \left(v_{k,j}^{O}\right)^2}} \tag{4}$$

x_{rk} is the measured operation similarity of OP_k between M_i and M_j and it is a part of the numerator in the above equation. It implies that even if these vectors equal to each other, according to the operation similarities, the operation vector similarity varies from 1 to 0. For the example of Figure 6, the measured operation vector similarity is $sim_{ov}(M_1, M_2) \approx \frac{3.32}{5.48} \approx 0.61$.

In [3], activity vectors contain not only the inclusion of activity but also the execution probability that is varied based on the routing patterns in business process models. Fundamentally, there are four basic routing patterns: sequential, parallel, selective, and iterative routings. Among them, the selective routing patterns such as OR-split and XOR-split decide execution probabilities of activities.

On the other hand, in cases of manufacturing processes, we empirically found that the manufacturing processes we investigate are composed of only sequential routings and parallel routings (less frequently), and they intrinsically have no selective and iterative routing patterns. Conventionally, manufacturing processes proceed only according to the predefined operations and their stationary execution orderings without any decision points that cause splitting control-flows patterns. However, in state-of-the-art manufacturing systems (e.g., flexible manufacturing systems), where process routing and machine assignments are flexible to react in case of changes, it is necessary to support the concept of the routing patterns not considered in this paper and supplement the definition of operation vector.

5.3.2. Transition Vector Similarity

The transition is the fundamental property of all classes of process models which is formally represented as a directed acyclic graph. The causality of tasks (including operations) in a process model is established based on transitions between the tasks, and it is a main structural property of process models including manufacturing processes. For example, the starting task in a process model precedes all other tasks including a succeeding task directly following it. Therefore, for manufacturing processes, causal relationships between operations are quantified and weighted through calculations of distance weights, and these are represented as a transition vector.

Let v_i^T be a transition vector of manufacturing process model M_i. v_i^T is a row vector containing $m \times m$ elements for all pairs of operations, where each element $v_{kl,i}^T$ represents a causal relationship between OP_k and OP_l, measured by the reverse of distance weight $d_{kl,i}^T$ between OP_k and OP_l.

$$v_i^T = \left[v_{11,i}^T, v_{12,i}^T, \ldots, v_{mm,i}^T\right] \tag{5}$$

$$v_{kl,i}^T = \frac{1}{d_{kl,i}^T} \tag{6}$$

The Cosine coefficient-based transition vector similarity $sim_{tv}(M_i, M_j)$ is measured by the following equation.

$$sim_{tv}(M_i, M_j) = \frac{\sum_{k=1}^{m} \sum_{l=1}^{m} v_{kl,i}^T\, v_{kl,j}^T}{\sqrt{\sum_{k=1}^{m} \sum_{l=1}^{m} \left(v_{kl,i}^T\right)^2}\ \sqrt{\sum_{k=1}^{m} \sum_{l=1}^{m} \left(v_{kl,j}^T\right)^2}} \tag{7}$$

Based on the above equation, the result of measuring transition vector similarity between M_1 and M_2 represented in Figure 7 is $sim_{tv}(M_1, M_2) \approx \frac{4.9}{7.16} \approx 0.684$.

The total similarity between process models M_i and M_j is measured by putting these two vector similarities $sim_{ov}(M_i, M_j)$ and $sim_{tv}(M_i, M_j)$ together and adding a balancing parameter $\alpha \in [0, 1]$ to blend them.

$$sim(M_i, M_j) = \alpha \cdot sim_{ov}(M_i, M_j) + (1 - \alpha) \cdot sim_{tv}(M_i, M_j) \tag{8}$$
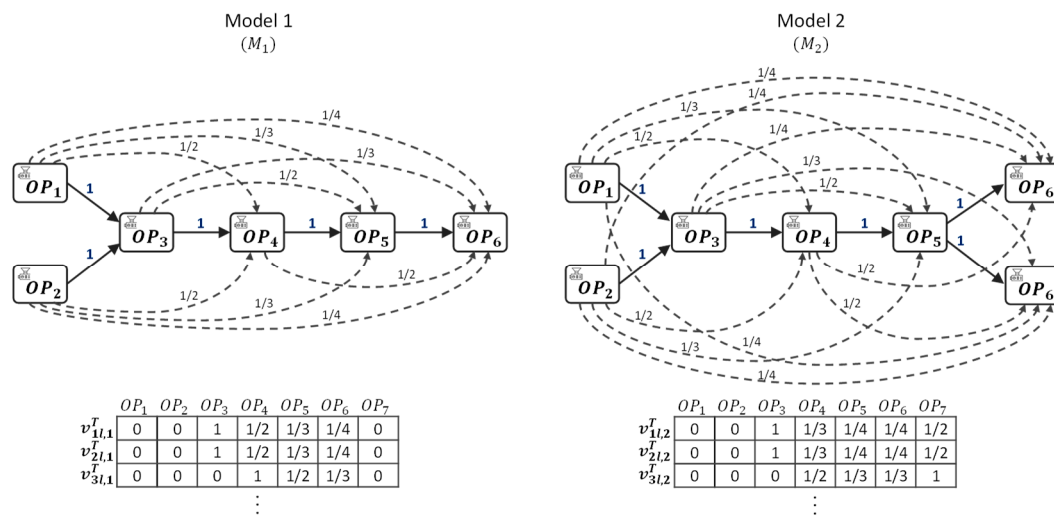

Figure 7. Representation of two transition vectors.

6. Similarity-Based Hierarchical Clustering

Based on the similarity measures, this section describes a hierarchical clustering algorithm of manufacturing process models. The hierarchical clustering is a useful technique that allows us to build a hierarchy of data clusters from similarities (or dissimilarities) between data. Hierarchical clustering methods generally fall into two types: the agglomerative (bottom-up) approach [24] and divisive (top-down) approach [25]. In this paper, we exploit an agglomerative approach, which is more common to the hierarchical clustering than the divisive approach, and we devise a clustering algorithm based on this approach.

6.1. Hierarchical Clustering Algorithm

The agglomerative clustering method builds a hierarchy from the n process models being clustered (see the algorithm below). Initially, each model is assigned to its own cluster (line 4–7) and then the clustering algorithm proceeds iteratively, at each step merging the two most similar clusters, continuing until all the models are merged into a single cluster (line 10–25).

Algorithm: Agglomerative Clustering using Process Similarity

Input: M, Set of manufacturing process models $(M_1, \ldots, M_n)$.
Output: H, Hierarchy of clusters $(\Lambda_1, \ldots, \Lambda_h)$.
1: **begin**
2: $l \leftarrow 1$;   //level of hierarchy
3: $\Lambda_l \leftarrow \varnothing$;
4: **for** $M_i \in \mathbf{M}$ //initialize a cluster set in level 1
5:    $\Lambda_l \leftarrow \varnothing$;
6:    $\Lambda_l \leftarrow \Lambda_l \cup M_i$;   //assign each model to its own cluster
7: **end for**
8: **H** $\leftarrow \varnothing$;
9: **H** $\leftarrow$ **H** $\cup \Lambda_l$;   //add the cluster set in level 1 to the hierarchy
10: **while** $(|\Lambda_l| > 1)$ //continue until all clusters are merged
11:   $\chi_m \leftarrow \varnothing$;   //a cluster for two clusters to be merged
12:   $max \leftarrow 0$;   //a variable for the global maximum similarity
13:   **for** $\chi_i \in \Lambda_l$ //find a maximum similarity between all pairs of clusters
14:      **for** $\chi_j \in \Lambda_l$, where $\chi_i \neq \chi_j$;
15:        **if** $sim_{max}^C(\chi_i, \chi_j) > max$ **then** //compare the global and local maximum similarities
16:          $\chi_m \leftarrow \varnothing$;
17:          $\chi_m \leftarrow \chi_i \cup \chi_j$;   //a set for the merged clusters
18:          $max \leftarrow sim_{max}^C(\chi_i, \chi_j)$;   //update the global maximal similarity
19:        **end if**
20:      **end for**
21:   **end for**
22:   $\Lambda_{l+1} \leftarrow \varnothing$;   //a cluster set in the next level
23:   $\Lambda_{l+1} \leftarrow ((\Lambda_l - \chi_i) - \chi_j) \cup \chi_m$, where $\chi_i, \chi_j \subseteq \chi_m$;   //subtract χ_i and χ_j, and add their merged cluster χ_m
24:   **H** $\leftarrow$ **H** $\cup \Lambda_{l+1}$;   //add the cluster set in level $l+1$ to the hierarchy
25:   $l \leftarrow l + 1$;   //increase the level of hierarchy
26: **return H**;

There are linkage-criteria, each of which determines a way to compute similarities between clusters. The complete-linkage method [26] takes the lowest similarity value (or farthest distance) between data included in the pair of clusters as a cluster similarity. Conversely, the single-linkage method [26] merges two clusters based on the highest similarity value (or closest distance) between pairs of clusters. These two representative linkage methods have slight differences in terms of time and space complexities. We accordingly use the single-linkage method that can easily be adapted to agglomerative clustering problems. Therefore, the manufacturing clustering algorithm calculates local maximum similarities, each of which indicates the maximum similarity observed from a pair of model clusters (line 13–18).

$$sim_{max}^C(\chi_i, \chi_j) = \max\{sim(M_a, M_b) : (M_a, M_b) \in \chi_i \times \chi_j\} \tag{9}$$

In the next step, the two model clusters having the global maximum similarity are selected to be merged. For example, let us three of model clusters χ_1, χ_2, and χ_3 at the level l. The algorithm calculates local maximum similarities of all pairs of clusters, $sim_{max}^C(\chi_1, \chi_2)$, $sim_{max}^C(\chi_1, \chi_3)$, and $sim_{max}^C(\chi_2, \chi_3)$. In case of the cluster pair of χ_1, χ_2, a local maximum similarity of this pair can be obtained by finding a maximal value of similarity among all possible pairs of process models, $\forall (M_a, M_b) : M_a \in \chi_1, M_b \in \chi_2$, included in χ_1 and χ_2. If the cluster pair of χ_2 and χ_3 has the global maximum similarity, these two clusters are joined into a new cluster χ_4 for the next level of the hierarchy. In this way, the algorithm merges the two closest clusters at each level and eventually completes the entire hierarchy of clusters.

6.2. Running Example

To confirm the applicability of the hierarchical clustering method for manufacturing process models, we present a running example with manufacturing process models. These models are selected models from real-life manufacturing processes for thermocouple probe products (see Figure 8) and obtained by the BPMN modeling phase from process charts, BOM and BOMO data. The details of the model examples are partially omitted or simplified for the security reasons. Specifically, the running example contains 8 process models ($\mathbf{M} = M_1, \ldots, M_8$), 15 operations ($\mathbf{OP} = OP_1, \ldots, OP_{15}$), and 57 components ($\mathbf{C} = C_1, \ldots, C_{57}$). According to the characteristics of the routing patterns of manufacturing processes, all the models have only sequential and parallel routing patterns in terms of control-flow.
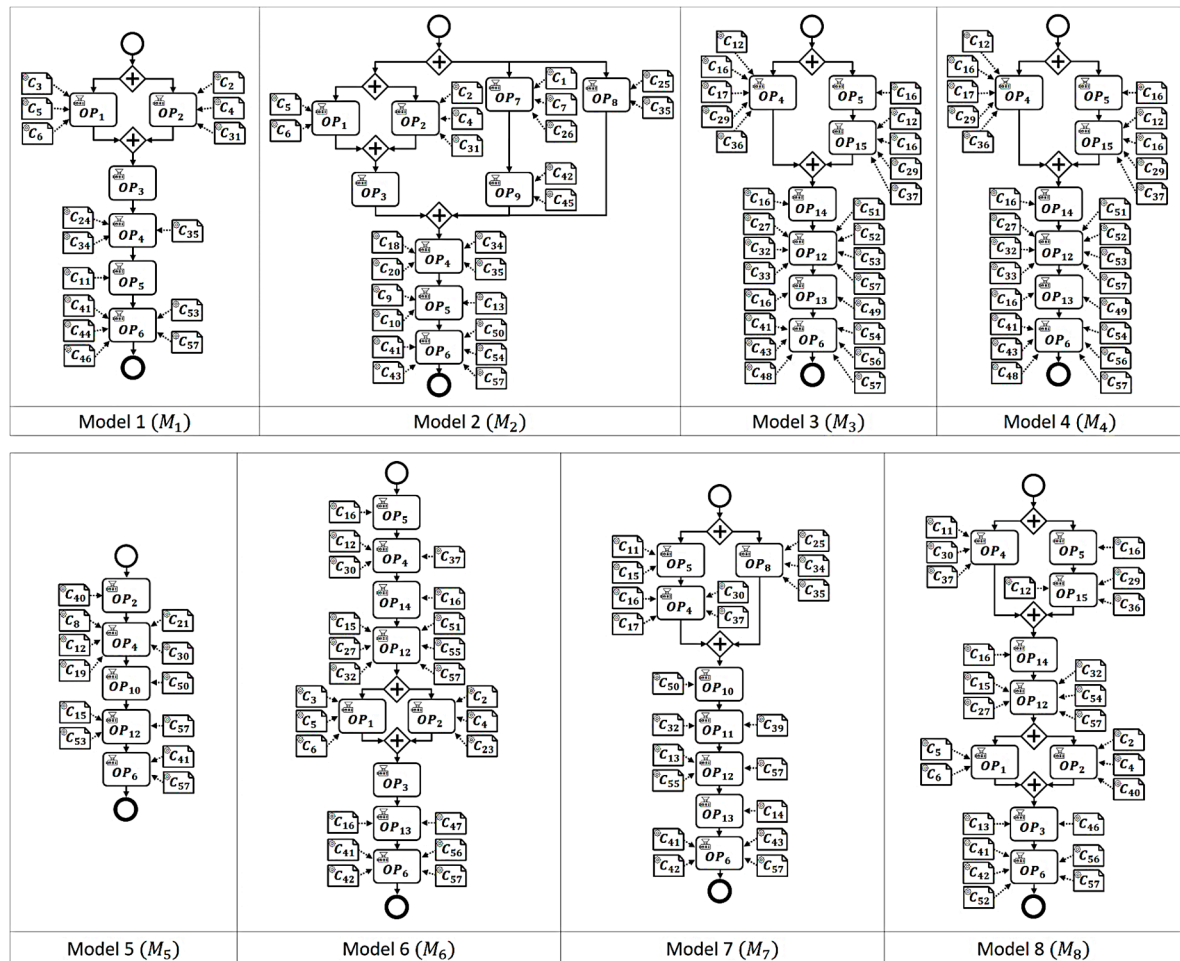

Figure 8. Manufacturing process models for the running example.

At the first step, an operation similarity matrix, $X = (x_{rk}) \in \mathbb{R}^{28 \times 15}$, between operations of the same type are calculated based on Equation (1). For example, the operation OP_4 is the operation of the same manufacturing task that is commonly included in all the process models. For the case of OP_4 in the model M_3, this operation is associated with the input components, C_{12}, C_{16}, C_{17}, C_{29}, and C_{36}. Due to the different operation–component associations from other process models, the measured operation similarities for OP_4^3 are

$$[0, 0, 1, 1, 0.11, 0.14, 0.29, 0]$$

Operation similarities affect the similarities between operation vectors, and Tables 4 and 5 represent the measured operation/transition vector similarities, respectively. The model pair (M_5, M_6) has the

highest operation vector similarity, $sim_{ov}(M_i, M_j) = 0.6709$, while the two most similar models on the transition vector similarity are M_3 and M_8 with the topmost value of 0.828. Ultimately, the measured total similarities with the balancing parameter $\alpha = 0.5$ are shown in Table 6. This result shows that the two most similar models are M_5 and M_6 with the similarity value of 0.6998.

Table 4. Measured operation vector similarities.

Process model	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
M_1	-	0.4518	0.1000	0.0343	0.3791	0.3289	0.0730	0.1134
M_2		-	0.0454	0.0720	0.2725	0.3056	0.0596	0.1375
M_3			-	0.0955	0.1055	0.0794	0.2313	0.3945
M_4				-	0.4152	0.3916	0.1197	0.1103
M_5					-	0.6709	0.1768	0.1515
M_6						-	0.1839	0.1229
M_7							-	0.2914
M_8								-

Table 5. Measured transition vector similarities.

Process model	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
M_1	-	0.7792	0.0154	0.0769	0.2873	0.3206	0.1276	0.0346
M_2		-	0.0120	0.0354	0.2239	0.2502	0.0995	0.0336
M_3			-	0.3553	0.1906	0.0415	0.4119	0.8280
M_4				-	0.4492	0.5320	0.1317	0.3138
M_5					-	0.7287	0.0812	0.2638
M_6						-	0.0982	0.0447
M_7							-	0.3414
M_8								-

Table 6. Total similarities with $\alpha = 0.5$.

Process model	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
M_1	-	0.6155	0.0577	0.0556	0.3332	0.3247	0.1003	0.0740
M_2		-	0.0287	0.0537	0.2482	0.2779	0.0796	0.0855
M_3			-	0.2254	0.1480	0.0604	0.3216	0.6113
M_4				-	0.4322	0.4618	0.1257	0.2120
M_5					-	0.6998	0.1290	0.2077
M_6						-	0.1410	0.0838
M_7							-	0.3164
M_8								-

Through the hierarchical clustering, we can expect to build a hierarchy of clusters of manufacturing process models using the measured similarities. Figure 9 shows the hierarchical clustering result for the running example, visualized as a heat map with dendrograms. As mentioned above, the hierarchical clustering method operates in the bottom-up agglomerative way and finds and merges a pair with the highest similarity (as Equation (9)) among all model pairs at each stage. In addition, by cutting the hierarchy at a certain level, an intended number of clusters can be obtained. This means that it is possible to determine the desired scale in managing manufacturing processes. For example, at the level 5 of the hierarchy, there are four clusters $\chi_1 = \{M_1, M_2\}$, $\chi_2 = \{M_3, M_7, M_8\}$, $\chi_3 = \{M_4\}$, and $\chi_4 = \{M_5, M_6\}$. As clusters χ_3 and χ_4 have the global maximum similarity $sim_{max}^C(\chi_3, \chi_4) = 0.4618$ resulted from the model pair ($M_4 \in \chi_3$, $M_6 \in \chi_4$), these two closest clusters are merged, resulting in a total of three clusters $\chi_1 = \{M_1, M_2\}$, $\chi_2 = \{M_3, M_7, M_8\}$, and $\chi_3 = \{M_4, M_5, M_6\}$ at the next level.
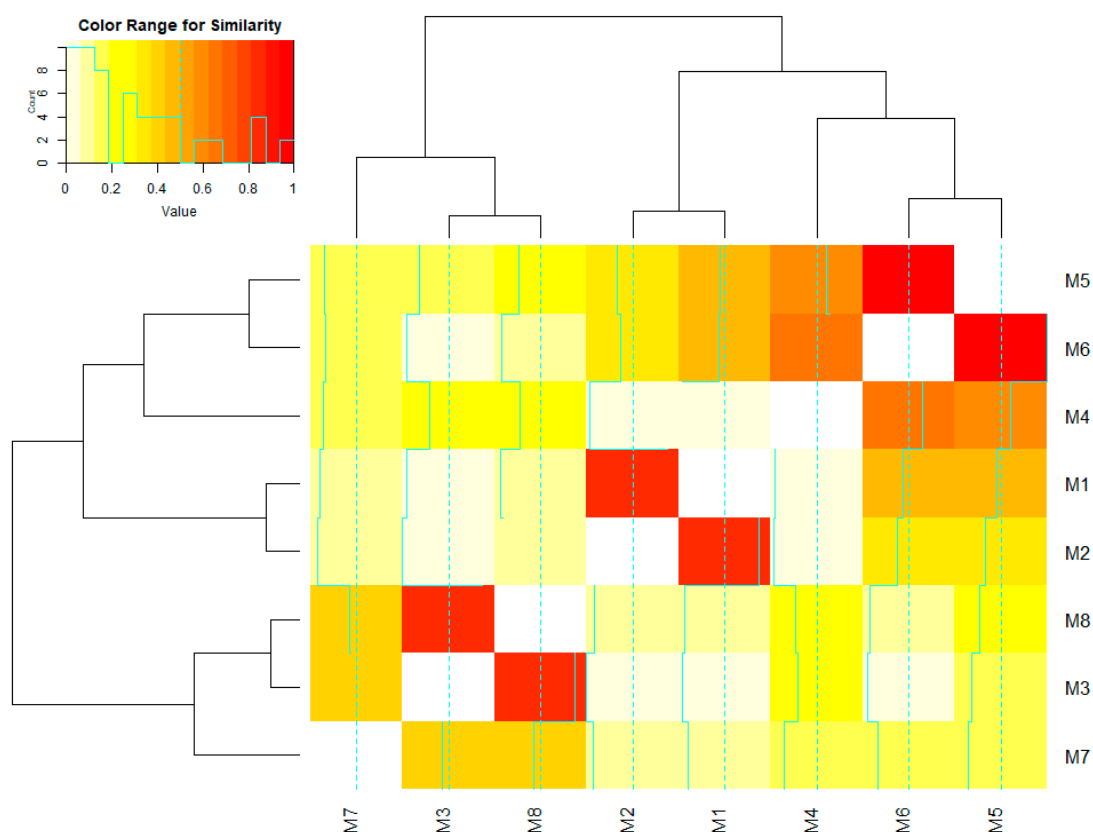
Figure 9. Hierarchical clustering result visualized as a heat map with dendrograms.

7. Conclusions

In the era of Industry 4.0, lots of manufacturing companies tend to have a vast array of manufacturing processes. With the attention of a need of manufacturing process management, this paper proposes a similarity-based hierarchical clustering method for manufacturing processes operated by manufacturing companies where the BPM methodology applied in. Specifically, the contributions of this paper are summarized as two-fold: (1) The manufacturing process modeling with the adaptation of the BPM approach. (2) The similarity measures for such manufacturing process models, and the hierarchical clustering based on the similarity. As the validation result of this work, the applicability of our clustering method was confirmed through the running example including real-life manufacturing processes for thermocouple products. In terms of manufacturing operation management, implications and possible applications are defined as follows:

BPM-compliant manufacturing process modeling—The process modeling standards generally provide visual notations and schema for storing textual information (e.g., BPMN). Thus, manufacturing process models that are visually standardized can help communications among related persons (e.g., process designers and stakeholders), at the same time, the textual information are interpreted by the BPM system to automatically execute manufacturing processes. Therefore, BPM-compliant manufacturing process models can be fully controlled in accordance with the BPM lifecycle. For example, process mining is an enabling technology to monitor status of entire manufacturing processes by analyzing event logs generated from the executions of process models.

Similarity-based hierarchical clustering of manufacturing process models—Clustering techniques aggregate similar objects into the same cluster. Thus, these facilitate group-level operations. For the manufacturing domain, operations in relevance with the production scheduling [27], resource planning and distribution can be performed effectively, centering on clustered manufacturing process groups. If there exist deficiencies such as bottleneck and low yield, which must be improved in the manufacturing process, the company may consider extending its reengineering task to the whole manufacturing

processes of the same cluster from the corresponding single process. For unexpected situations, especially, the alternation operation of a manufacturing process or related resources (e.g., machines, components, and human resources) is also one viable application which is activated by manufacturing process clustering. Assuming a certain company operating a large-scale set of manufacturing processes that is the target environment of this paper, our clustering method can support in designing manufacturing processes for new products. By searching based on the similarity measures, process designers can find a matched process or partial information, valuable to be referred or reused.

In conclusion, we believe that our method will support manufacturing companies so as to design and manage a vast amount of manufacturing processes at a coarser level. However, in order for our method to be more realistic, the following points should be supplemented in a follow-up study. First, the BPMN modeling is insufficient to capture the other important characteristics of the manufacturing process. Machines and human resources are also important entities that steer the manufacturing process, therefore the modeling functionality supporting them should also be added. In addition, we need to review other similarity techniques (e.g., graph similarity) and apply it to make our method more reasonable and effective. Second, this work does not address the evaluation of how the proposed clustering method can contribute to quantitative and qualitative positive effects on manufacturing operation management and engineering problems. Therefore, our future works will cover the following subjects:

- Studying process modeling and similarity techniques for better capturing the characteristics of manufacturing processes.
- Conducting a case study applying the proposed clustering methods to a large set of real-life manufacturing processes and investigating its effects.
- Design and implementation of a system supporting decision-making in various engineering problems based on clustering results.

We are looking into implanting the BPM approach and systems into the manufacturing company of thermocouple products. In a broad sense, the ultimate objective of our future works is to establish realistic applications that provide an effective means supporting BPM-based manufacturing operation management.

Author Contributions: H.A. and T.C. did the initial design and analysis on the research; H.A. performed experiments; H.A. and T.C. analyzed the data; H.A. prepared the bulk of the manuscript and wrote the draft of the paper; while T.C. contributed write-up for the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hwang, G.; Kang, S.; Dweekat, A.; Park, J.; Chang, T.-W. An IoT data anomaly response model for smart factory performance measurement. *Int. J. Ind. Eng. Theory Appl. Pract.* **2018**, *25*, 702–718.
2. Ahn, H.; Chang, T.-W. Measuring Similarity for Manufacturing Process Models. In Proceedings of the IFIP International Conference on Advances in Production Management Systems, Seoul, Korea, 27–30 August 2018; pp. 223–231.
3. Jung, J.-Y.; Bae, J.; Liu, L. Hierarchical Clustering of Business Process Models. *Int. J. Innov. Comput. Inf. Control* **2009**, *5*, 1349–4198.
4. Dijkman, R.; Dumas, M.; Garcia-Banuelos, L. Graph Matching Algorithms for Business Process Model Similarity Search. In Proceedings of the International Conference on Business Process Management, Ulm, Germany, 8–10 September 2009; pp. 48–63.
5. Ivan, A.; Akkiraju, R. Discovering Business Process Similarities: An Empirical Study with SAP Best Practice Business Processes. In Proceedings of the International Conference on Service-Oriented Computing, San Francisco, CA, USA, 7–10 December 2010; pp. 515–526.

6. Kunze, M.; Weske, M. Metric Trees for Efficient Similarity Search in Large Process Model Repositories. In Proceedings of the International Conference on Business Process Management, Hoboken, NJ, USA, 13–16 September 2010; pp. 535–546.

7. Schoknecht, A.; Thaler, T.; Fettke, P.; Oberweis, A.; Laue, R. Similarity of Business Process Models: A State-of-the-Art Analysis. *ACM Comput. Surv.* **2017**, *50*, 52. [CrossRef]

8. Gupta, T.; Seifoddini, H. Production Data Based Similarity Coefficient for Machine-Component Grouping Decisions in the Design of a Cellular Manufacturing System. *Int. J. Prod. Res.* **1990**, *28*, 1247–1269. [CrossRef]

9. Dimopoulos, C.; Mort, N. A Hierarchical Clustering Methodology Based on Genetic Programming for the Solution of Simple Cell-Formation Problems. *Int. J. Prod. Res.* **2001**, *39*, 1–19. [CrossRef]

10. Park, S.; Suresh, N.C. Performance of Fuzzy ART Neural Network and Hierarchical Clustering for Part-Machine Grouping Based on Operation Sequences. *Int. J. Prod. Res.* **2003**, *41*, 3185–3216. [CrossRef]

11. Li, B.M.; Xie, S.Q. Module Partition for 3D CAD Assembly Models: A Hierarchical Clustering Method Based on Component Dependencies. *Int. J. Prod. Res.* **2015**, *53*, 5225–5240. [CrossRef]

12. Daie, P.; Li, S. Managing Product Variety through Configuration of Pre-Assembled Vanilla Boxes using Hierarchical Clustering. *Int. J. Prod. Res.* **2016**, *54*, 5468–5479. [CrossRef]

13. Etgar, R.; Gelbard, R.; Cohen, Y. Presenting the Several-Release-Problem and Its Cluster-Based Solution Acceleration. *Int. J. Prod. Res.* **2017**. Available online: https://doi.org/10.1080/00207543.2017.1404657 (accessed on 2 May 2019). [CrossRef]

14. Weske, M. *Business Process Management: Concepts, Languages, Architectures*; Springer: Berlin/Heidelberg, Germany, 2012.

15. Stefanini, A.; Aloini, D.; Benevento, E.; Dulmin, R.; Mininno, V. Performance Analysis in Emergency Departments: A Data-driven Approach. *Meas. Bus. Excell.* **2018**, *22*, 130–145. [CrossRef]

16. Van der Aalst, W.M.P.; Reijers, H.A.; Weijters, A.J.M.M.; van Dongen, B.F.; Alves de Medeiros, A.K.; Song, M.; Verbeek, H.M.W. Business Process Mining: An Industrial Application. *Inf. Syst.* **2007**, *32*, 713–732. [CrossRef]

17. Zerbino, P.; Aloini, D.; Dulmin, R.; Mininno, V. Process-mining-enabled audit of Information Systems: Methodology and an application. *Expert Syst. Appl.* **2018**, *110*, 80–92. [CrossRef]

18. Sungur, C.T. Extending BPMN for Wireless Sensor Networks. Master's Dissertation, University of Stuttgart, Stuttgart, Germany, 2013.

19. Ruiz, F.; Garcia, F.; Calahorra, L.; Llorente, C.; Goncalves, L.; Daniel, C.; Blobel, B. Business Process Modeling in Healthcare. *Stud. Health Technol. Inform.* **2012**, *179*, 75–87. [PubMed]

20. Zor, S.; Leymann, F.; Schumm, D. A Proposal of BPMN Extensions for the Manufacturing Domain. In Proceedings of the 44th CIRP International Conference on Manufacturing Systems, Madison, WI, USA, 31 May–3 June 2011.

21. Bocciarelli, P.; D'Ambrogio, A.; Giglio, A.; Paglia, E. A BPMN Extension for Modeling Cyber-Physical-Production-Systems in the Context of Industry 4.0. In Proceedings of the 14th IEEE International Conference on Networking, Sensing and Control, Calabria, Italy, 16–18 May 2017; pp. 599–604.

22. KSA (Korean Standards Association). KS-A-3002 Standard: Graphical Symbols for Process Chart. Available online: https://www.kssn.net/en/search/stddetail.do?itemNo=K001010103692 (accessed on 2 May 2019).

23. Jiao, J.; Tseng, M.M.; Ma, Q.; Zou, Y. Generic Bill-of-Materials-and-Operations for High-Variety Production Management. *Concur. Eng.* **2000**, *8*, 297–321. [CrossRef]

24. Murtagh, F.; Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classif.* **2014**, *31*, 274–295. [CrossRef]

25. Savaresi, S.M.; Boley, D.L. On the Performance of Bisecting K-means and PDDP. In Proceedings of the 2001 SIAM International Conference on Data Mining, Chicago, IL, USA, 5–7 April 2001; pp. 1–14.

26. Wilks, D.S. Cluster Analysis. *Int. Geophysics* **2011**, *100*, 603–616.

27. De Matta, R. Scheduling a Manufacturing Process with Restrictions on Resource Availability. *Int. J. Prod. Res.* **2017**, *56*, 6412–6429. [CrossRef]